www.jusletter-it.eu

Rolf H. Weber / Dominic Oertly

Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics?

The analysis of large volumes of structured and unstructured data from manifold sources and the establishment of correlations between data in their processing leads to the risk that it becomes possible to draw conclusions from non-personal data to identifiable persons. Big data analytics can cause the de-anonymization of data. Even if technial procedures exist that make it difficult to establish a relation between anonymized data and persons, it must be acknowledged that these procedures are technically often complicated to apply. In order to achieve a risk minimization, an interdisciplinary cooperation between organization, technique and law must be realized.

Category: Articles
Field of law: Big Data, Open Data & Open Government; Data Protection
Region: Switzerland

Citation: Rolf H. Weber / Dominic Oertly, Aushöhlung des Datenschutzes durch De-Anonymisierung bei Big Data Analytics?, in: Jusletter IT 21 May 2015

Inhaltsübersicht

1 Big Data — Grundlagen und Problemstellung
1.1 Wesensmerkmale
1.2 Rechtliche Einbettung und Problemfelder
2 Anonymisierung und De-Anonymisierung von Daten
2.1 Vorgang der Anonymisierung
2.2 Risiko der De-Anonymisierung
2.3 Rechtliche Vorgaben in Europa?
3 «Management» von Big Data und Lösungsansätze
3.1 Big Data Governance
3.2 Risikominderung bei der Datenbearbeitung
4 Ausblick

1 Big Data — Grundlagen und Problemstellung

«*Big Data is high-volume, high-velocity and high- variety information assets that demand cost- effective, innovative forms of information processing for enhanced insight and decision making.*»[1]

1.1 Wesensmerkmale

[Rz 1] Big Data bezieht sich nicht nur auf den Einsatz grosser Mengen an strukturierten und unstrukturierten Daten, sondern auch auf deren Analyse durch sog. «Big Data Analytics». Mit Hilfe dieser Analysetools ist eine fortlaufende Analyse sämtlicher Daten möglich.[2] Im Rahmen der Verarbeitung findet eine immer stärkere Korrelation unter Daten im Lichte einer unlimitierten Menge an Informationen statt: Standen bisher das Data Mining/Data Warehousing sowie die Beantwortung von durch das Unternehmen vordefinierten Fragen im Vordergrund,[3] sind moderne Analysetools nun in der Lage, auch unstrukturierte Daten miteinzubeziehen.[4]

[Rz 2] Big Data zielt darauf ab, umfassende Bestände an Daten, die aus unterschiedlichen Quellen stammen, in Hochleistungsdatenbanken zu sammeln und auszuwerten.[5] Die Datenberge werden automatisiert und nach unbekannten Korrelationen durchsucht, dennoch stehen die Resultate innert Sekunden zur Verfügung. Weil sich die zugrunde liegenden lernfähigen Algorithmen von selbst optimieren,[6] hat deren Einsatz eine ständige Effizienzsteigerung zur Folge. Schliesslich sollen Big Data Analytics dazu dienen, neben den bereits heute gestellten Fragen auch neue Muster zu erkennen.[7]

---

1 Vgl. die Definition von Gartner Inc., dem nach eigenen Angaben globalen Leader für Marktforschung und Beratung im Bereich der Informationstechnologie, http://www.gartner.com/it-glossary/big-data/ (alle Internetquellen wurden zuletzt am 4. Mai 2015 überprüft).

2 Bruno Baeriswyl, Big Data zwischen Anonymisierung und Re-Individualisierung, in: Weber/Thouvenin (Hrsg.), Big Data und Datenschutz — Gegenseitige Herausforderungen, ZIK Band 59, Zürich 2014, 46.

3 Bereits bei der Filterung und Auswahl der Daten müsste ein Unternehmen genau wissen, welche Fragestellungen die Analyse beantworten sollte, vgl. dazu Georg Polzer, Big Data — eine Einführung, in: digma 2013, 6.

4 Baeriswyl (Fn 2), 46, 48.

5 Rolf H. Weber, Big Data: Sprengkörper des Datenschutzrechts, in: Jusletter IT 11. Dezember 2013, N 1; Christoph Zieger/Nikolas Smirra, Fallstricke bei Big Data-Anwendungen, MMR 2013, 418; Ira S. Rubinstein, Big Data: The End of Privacy or a New Beginning?, International Data Privacy Law 2013, Vol. 3, No. 2, 74, 77.

6 Vgl. Rubinstein (Fn 5), 76, wonach dieses Phänomen als «Data mining on steroids» zu bezeichnen sei; Baeriswyl (Fn 2), 48.

7 Bruno Baeriswyl, «Big Data» ohne Datenschutz-Leitplanken, in: digma 2013, 14.

[Rz 3] Bei der umfassenden Datenverfügbarkeit geht es somit nicht mehr einfach nur um die Frage nach dem «warum», sondern um das «was»; somit führen Datenanalysen durch Big Data weg von der herkömmlichen Kausalitätsdiskussion.[8] Die Entwicklung bewegt sich von einem Konzept der Kausalität zu einem Korrelationskonzept hin.[9] Die Tatsache, dass das Konzept der Korrelation und nicht mehr jenes der Kausalität massgebend wird, dürfte für die traditionelle Rechtswissenschaft eine gewisse Anpassungsnotwendigkeit nach sich ziehen.

[Rz 4] Die quantitative Veränderung der Datenverarbeitung hat ohne Zweifel auch qualitative Veränderungen zur Folge: Werden mehr Daten verarbeitet, ist es sehr wahrscheinlich, dass zusätzliche datenschutzrechtliche Risiken auftreten.[10] Weil Big Data durch die modernen Datenzugriffs- und Datenauswertungsmethoden eine neue Form der Wertschöpfung etabliert hat, beziehen sich die Risiken neben der üblichen Erstverwendung vermehrt auch auf die Zweitverwendung der Daten.[11] Fraglich bleibt, inwieweit die relativ traditionelle Datenschutzgesetzgebung auf diese Herausforderung zu reagieren vermag.

1.2 Rechtliche Einbettung und Problemfelder

[Rz 5] Ausgangspunkt für die datenschutzrechtliche Beurteilung von Big Data ist die Frage, ob das hiesige Datenschutzgesetz (DSG)[12] Anwendung findet. Das DSG regelt den Schutz von Personendaten und die Einhaltung verschiedener Bearbeitungsprinzipien im Falle einer Datensammlung. Weisen gewisse Daten keinen Personenbezug auf und lässt sich ein entsprechender Bezug auch nicht herstellen, greift das DSG nicht, weil ein relevantes Schutzsubjekt fehlt.[13] Wenn es um reine Sachdaten geht, kommt das DSG somit grundsätzlich nicht zur Anwendung.

[Rz 6] Die Definition der Personendaten (Art. 3 lit. a DSG) sieht vor, dass ein Bezug zu einer bestimmten oder bestimmbaren Person bestehen muss. Ergibt sich die Identität einer Person aus der Information selbst, liegt Bestimmtheit im Sinne des Gesetzes vor.[14] Bestimmbar ist eine Person, wenn sich ein Personenbezug ohne unverhältnismässigen Aufwand erstellen lässt und auch damit gerechnet werden muss, dass dieser potentiell erfolgt.[15] Ein Aufwand erscheint als unverhältnismässig, wenn nach allgemeiner Lebenserfahrung nicht davon auszugehen ist, dass ein Datenbearbeiter diesen auf sich nehmen wird.[16] Angesichts der informationstechnischen Möglichkeiten ist der Aufwand, der für die Bestimmung von Personen erforderlich ist, im Rahmen von Big Data in den meisten Fällen verhältnismässig.[17] Somit drängt sich die Frage auf, wie mit Ergebnissen umzugehen ist, die unbewusst einen Personenbezug herstellen lassen.[18]

---

8 Weber (Fn 5), N 7; Viktor Mayer-Schönberger/Kenneth Cukier, Big Data, A Revolution, New York 2013, 12 ff.

9 Mayer-Schönberger/Cukier (Fn 8), 53, 63.

10 Weber (Fn 5), N 5.

11 Mayer-Schönberger/Cukier (Fn 8), 153; Weber (Fn 5), N 5.

12 Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992 (SR 235.1).

13 Statt vieler vgl. Rolf H. Weber, Datenschutzrecht vor neuen Herausforderungen, ZIK Band 13, Zürich 2000, 123 f.

14 Gabor P. Blechta, in: Maurer-Lambrou/Blechta (Hrsg.), Basler Kommentar (BSK) zum Datenschutzgesetz und Öffentlichkeitsgesetz, 3. Aufl., Basel 2014, Art. 3 DSG N 9.

15 Statt vieler vgl. Baeriswyl (Fn 2), 49.

16 BSK-Blechta (Fn 14), Art. 3 DSG N 11.

17 Baeriswyl (Fn 2), 50.

18 Baeriswyl (Fn 2), 50, 53.

[Rz 7] In diesem Kontext sind beim Vorgang der Datenbearbeitung insbesondere drei Faktoren zu berücksichtigen:[19]

(i) Dateninput: Daten sind strukturiert oder unstrukturiert und stammen aus unterschiedlichen Quellen, wie z.B. Instrumenten, Sensoren, click streams oder Internetauftritten. *(ii) Datenprozessierung:* In der Realität besteht eine Vielzahl an Plattformen, auf denen die Daten erscheinen können, welche schliesslich eine Datenbearbeitung ermöglichen. *(iii) Datenoutpout:* Beim Datenoutput steht die Frage im Zentrum, ob bestimmte Daten lediglich für einen internen Zweck zur Verwendung stehen sollen oder ob ebenfalls eine Nutzung für externe Zwecke denkbar ist.

[Rz 8] Liegen Personendaten vor, sind insbesondere die Grundsätze nach Art. 4—7 DSG zu berücksichtigen. Sobald die Anonymisierung von Daten erfolgt ist, gibt es aus Sicht des Datenschutzes indessen grundsätzlich kein Problem mehr, weil die entstandenen Sachdaten nicht in den Anwendungsbereich des DSG fallen.[20] Die Frage ist jedoch, ob die Anonymisierung von Daten nicht das Risiko der De-Anonymisierung nach sich zieht (bzw. dieses Risiko nicht ausschliesst) und damit die Datenbearbeitung nach einem solchen Vorgang wieder DSG-relevant wird.

2 Anonymisierung und De-Anonymisierung von Daten

2.1 Vorgang der Anonymisierung

[Rz 9] Daten gelten als anonymisiert, wenn ein Bezug zu einer Person nicht (mehr) möglich ist. Damit nicht nur eine Pseudo-Anonymisierung vorliegt, muss dieser Vorgang irreversibel sein.[21] Bei pseudo-anonymisierten Daten ist sowohl die Verknüpfung unterschiedlicher Datenbestände als auch das Herausgreifen einzelner betroffener Personen weiterhin möglich, weshalb sie mit anonymisierten Daten nicht gleichzusetzen sind.[22] Wie ein Ereignis bei America On Line (AOL) exemplarisch zeigte, sind pseudo-anonymisierte Daten geeignet, eine Identifizierung zu ermöglichen.[23]

[Rz 10] Weil durch den Prozess der Anonymisierung eine Datenbearbeitung stattfindet, sind die Grundsätze des DSG anzuwenden. Der Umgang mit dem Ergebnis der Anonymisierung — den anonymisierten Daten — ist hingegen nicht mehr datenschutzrelevant.[24] Während also die Erfassung der Daten datenschutzrechtliche Implikationen zeitigt, ist die Datenverwendung nicht weiter DSG-relevant. Für die Bearbeitung von personenbezogenen Daten zu Zwecken, die nicht

---

19 Vgl. dazu vertiefend Richard Kemp, Legal aspects of managing Big Data, Computer Law & Security Review 30 (2014), 489 f.

20 Vgl. vorne Ziff. 1.2.

21 Günther Karjoth, Sind anonymisierte Daten anonym genug?, in: digma 2008, 18 ff.; Baeriswyl (Fn 2), 50.

22 Artikel-29-Datenschutzgruppe der Europäischen Union, Stellungnahme 5/2014 zu Anonymisierungstechniken, angenommen am 10. April 2014, 12, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf. Diese Datenschutzgruppe vereinigt eine Vielzahl von Experten des Datenschutzes und gibt relativ oft Empfehlungen heraus.

23 Im Jahr 2006 hat AOL eine Datenbank mit 20 Millionen Suchwörtern publiziert. Als einzige Massnahme für den Datenschutz hat AOL die Nutzer-ID durch einen numerischen Wert ersetzt; in der Folge wurde eine öffentliche Identifizierung einiger Nutzer möglich, was in der Öffentlichkeit für Aufsehen sorgte. Die (Pseudo-)Anonymisierung war somit ungenügend. Vgl. dazu Artikel-29-Datenschutzgruppe (Fn 22), 12.

24 Baeriswyl (Fn 2), 50.

personenbezogen sind, liegt somit in der Regel ein Rechtfertigungsgrund vor.[25]

[Rz 11] Eine Anonymisierung von Personendaten garantiert indessen grundsätzlich nicht die Anonymität bzw. den Ausschluss einer De-Anonymisierung. In diesem Kontext weist die Artikel-29-Datenschutzgruppe der Europäischen Union ausdrücklich darauf hin, dass die technischen Verfahren zur Anonymisierung und De-Anonymisierung Gegenstand laufender Forschung sind; es habe sich bereits mehrfach gezeigt, dass (noch) keine Technik ohne Mängel sei.[26]

[Rz 12] Bei den technischen Verfahren zur Anonymisierung ist zwischen der Randomisierung und der Generalisierung zu unterscheiden:

(i) Unter Randomisierung sind Techniken zu verstehen, welche Daten insoweit verfälschen, als eine unmittelbare Verbindung zwischen den betroffenen Personen und ihren Daten entfernt wird. Weil bei diesem Anonymisierungsansatz jeder Datensatz nach wie vor eine einzige betroffene Person zum Gegenstand hat, besteht die Einzigartigkeit der einzelnen Datensätze fort.[27] Als Unterkategorien der Randomisierung sind die stochastische Überlagerung, die Vertauschung und die Differential Privacy zu nennen; sie beinhalten die nachfolgenden Merkmale:

- *Stochastische Überlagerung*: Mit Hilfe dieser Anonymisierungstechnik werden Merkmale in einem Datenbestand so verändert, dass sie weniger genau sind, obwohl die generelle Verteilung erhalten bleibt; im Zuge der Verarbeitung eines Datenbestandes entsteht der Eindruck präziser Werte, welcher jedoch in einem gewissen Masse täuschend ist.[28]
- *Vertauschung:* Die Anonymisierungstechnik der Vertauschung basiert auf der Ersetzung von Merkmalswerten, damit eine künstliche Verknüpfung mit anderen Personen stattfindet. Ein solches Vorgehen kann sinnvoll sein, wenn die exakte Verteilung eines jeden Merkmals im Datenbestand aufrechtzuerhalten ist.[29]
- *Differential Privacy:* Das Konzept der Differential Privacy ist heranzuziehen, wenn die für die Bearbeitung zuständige Stelle anonymisierte Ansichten eines Datenbestandes generiert und zugleich eine Kopie der Originaldaten aufbewahrt.[30]

(ii) Im Gegensatz zur Randomisierung bezweckt die Generalisierung die Verallgemeinerung der Merkmale betroffener Personen durch die Veränderung von Grössenskalen und -ordnungen. Diese Form der Anonymisierung kann zwar das Herausgreifen einzelner Personen verhindern, sie ermöglicht aber nicht zwingend eine effektive Anonymisierung.[31] Die k-Anonymität und die i-Diversität als Teil der Generalisierung vermögen für den Schutz der Offenlegung von Attributen und den Schutz der Verknüpfung von Identitäten zu sorgen. Es geht jeweils darum, dass mit Blick auf sog. k-1 Faktoren die entsprechende Person nicht mehr erkennbar ist.[32]

---

25 Vgl. dazu Art. 13 Abs. 2 lit. e DSG und für weitere Einzelheiten Baeriswyl (Fn 2), 50 f.

26 Artikel-29-Datenschutzgruppe (Fn 22), 13.

27 Eingehender dazu Artikel-29-Datenschutzgruppe (Fn 22), 14.

28 Artikel-29-Datenschutzgruppe (Fn 22), 14.

29 Artikel-29-Datenschutzgruppe (Fn 22), 16.

30 Cynthia Dwork, «Differential Privacy», in: Automata, languages and programming, Berlin/Heidelberg 2006, 1 ff.; Artikel-29-Datenschutzgruppe, (Fn 22), 17 f.

31 Vgl. zum Ganzen Artikel-29-Datenschutzgruppe (Fn 22), 19.

32 Karjoth (Fn 21), 20 ff.

- *k-Anonymität:* Die k-Anonymität zielt darauf ab, das Herausgreifen einer betroffenen Person zu verhindern, indem diese mit mindestens k anderen Personen zusammengefasst wird. Um dies zu erreichen, werden die Merkmalswerte in einem Masse verallgemeinert, dass alle k-Personen denselben Merkmalswert aufweisen.[33]
- *i-Diversität:*[34] Das Konzept der i-Diversität erweitert die k-Anonymität; dabei sollen keine gleich ablaufenden Angriffe durch Inferenztechniken mehr möglich sein, indem dafür gesorgt wird, dass einzelne Merkmale in jeder Äquivalenzklasse mindestens i unterschiedliche Werte beinhalten. Eine Konkretisierung der i-Diversität stellt die t-closeness dar, die Äquivalenzklassen bildet, welche der ursprünglichen Verteilung der Werte ähnlich sind.[35]

[Rz 13] Ob die dargelegten technischen Verfahren der Randomisierung und der Generalisierung die Risiken einer De-Anonymisierung der Daten tatsächlich ausschliessen, erscheint indessen als (zumindest) fraglich.

2.2 Risiko der De-Anonymisierung

[Rz 14] Die De-Anonymisierung von Daten meint das Rückgängigmachen einer Anonymisierung.[36] Materiell geht es um einen Prozess, den es datenschutzrechtlich eigentlich nicht gibt (oder nicht geben sollte).[37] Die Anonymisierung sollte ja irreversibel sein, so dass kein Personenbezug mehr herstellbar ist.

[Rz 15] Bei der Analyse von anonymisierten Daten ist die Menge entscheidend. Je grösser die Datenmenge ist, desto höher wird auch die Wahrscheinlichkeit, dass Daten einer bestimmten Person zugeordnet werden können.[38] Eine bekannte Studie aus den Vereinigten Staaten belegt, dass drei (relativ einfache) demographische Merkmale, nämlich Geschlecht, Geburtsdatum und fünf-stellige Postleitzahl — je nach Lokalisierung — es ermöglichen, zwischen 61% und 87% der amerikanischen Bevölkerung eindeutig zu identifizieren.[39]

[Rz 16] Zudem haben Studien im Bereich der Forschung gezeigt, dass anonyme Gen-Sequenzen, welche sich auf öffentlich zugänglichen Forschungs-Datenbanken befinden, durch Kombination mit wenigen anderen Daten eine «De-Anonymisierung» erlauben.[40] Auch gibt es Untersuchungen, die zeigen, dass man aus gewissen Sachdaten relativ gut auf den Verlauf von Krankheiten schliessen kann;[41] somit müssen Big Data Analytics nicht per se negativ sein. Abgesehen vom

---

[33] Für weitere technische Hinweise vgl. Latanya Sweeney, k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuziness and Knowledge-based Systems, 10 (5), 2002, 557 ff.; Artikel-29-Datenschutzgruppe (Fn 22), 19.

[34] In der Literatur wird teilweise auch der Begriff «l-Diversität» verwendet.

[35] Für weitere technische Hinweise vgl. Wang Pingshui/Wang Jiandong, L-Diversity Algorithm for Incremental Data Release, Appl. Math. Inf. Sci. 7, No. 5, 2013, 2055 ff.; Artikel-29-Datenschutzgruppe (Fn 22), 22.

[36] Dieser Vorgang ist auch als «Re-Individualisierung» bekannt, vgl. dazu Baeriswyl (Fn 2), 51.

[37] Baeriswyl (Fn 2), 51.

[38] Statt vieler vgl. Baeriswyl (Fn 2), 52.

[39] Vgl. Karjoth (Fn 21), 18 ff. für diese und weitere Studien sowie Weber (Fn 5), N 13.

[40] Melissa Gymrek/Amy L. McGuire/David Golan/Eran Halperin/Yaniv Erlich, Identifying Personal Genomes by Surname Inference, Science No. 339/6117, Januar 2013, 321 ff.

[41] Dabei ist z.B. an die Krankheit Ebola zu denken, vgl. dazu http://www.cnbc.com/id/102049616.

Gesundheitsbereich gibt es weitere praktische Beispiele (z.B. die Re-Individualisierung von anonymisierten Steuerdaten), welche die Risiken einer De-Anonymisierung offenbaren.[42]

[Rz 17] Wie festgestellt gibt es verschiedenste Bereiche, in welchen die Anonymisierung instabil ist,[43] wenn Big Data Analytics betrieben wird. Probleme können insbesondere auftauchen in Bezug auf Personendaten durch den sog. «Zufallsfund», mit Blick auf die personenbezogene Zwecksetzung und im Kontext der Bekanntgabe anonymer Daten.[44] Angesichts der Analysedaten und der Möglichkeit von Data Sharing (Austausch von Daten) besteht in vielen Bereichen ein nicht zu unterschätzendes Risiko der De-Anonymisierung.

[Rz 18] Die Bekanntgabe anonymer Daten fällt nicht unter das DSG.[45] Ist hingegen mit einer De-Anonymisierung zu rechnen, handelt es sich nicht um anonyme Daten im Sinne des Gesetzes. Im Kontext der Big Data Analytics ist indessen mit einer immer höheren Wahrscheinlichkeit damit zu rechnen, dass es zu einer De-Anonymisierung von anonymen Daten kommt.[46] Um die Wahrscheinlichkeit möglichst tief zu halten, ist es notwendig, die im konkreten Fall passende Anonymisierungstechnik zu wählen. Im Zusammenhang mit den technischen Verfahren der Anonymisierung sind namentlich drei Risiken von zentraler Bedeutung:[47]

- *Herausgreifen von Datensätzen:* Das Risiko des Herausgreifens besteht darin, dass sich in einem Datenbestand bestimmte Datensätze isolieren lassen, mit der Folge, dass die Identifizierung einer Person potentiell möglich wird.
- *Inferenz:* Dieses Risiko bezieht sich auf die Möglichkeit, den Wert eines Merkmals mit einer Wahrscheinlichkeitsanalyse von den Werten einer Reihe mit anderen Merkmalen abzuleiten.
- *Verknüpfbarkeit:* Dieses Risiko berücksichtigt die Verknüpfung mindestens zweier Datensätze (in derselben Datenbank oder in zwei verschiedenen Datenbanken), welche dieselbe Person oder Personengruppe betreffen. Ist ein Interessent fähig — z.B. durch eine Korrelationsanalyse — festzustellen, dass zwei Datensätze dieselbe Personengruppe betreffen, ohne aber einzelne Personen in dieser Gruppe herauszugreifen, bietet die entsprechende Technik keinen Schutz vor Verknüpfbarkeit.

[Rz 19] Die Risiken einer De-Anonymisierung unterscheiden sich je nach gewählter Anonymisierungstechnik. Weil eine diesbezüglich vertiefende Analyse den Rahmen des vorliegenden Beitrages sprengen würde, ist auf die Ausführungen der Artikel-29-Datenschutzgruppe der Europäischen Union zu verweisen.[48]

2.3 Rechtliche Vorgaben in Europa?

[Rz 20] Soweit ersichtlich, gibt es derzeit keine Gesetzgebung, der es gelungen ist, die Phänomene von Big Data und deren Risiken in den Griff zu bekommen. Diese Tatsache soll jedoch

[42] Steve Saxby/Alison Knight/Henry Pearce, Piercing the Anonymity Veil: Re-identification risk and the UK Transparency Agenda, in: Kierkegaard (Hrsg.), Information Ethics and Security: Future of International World Time, Kopenhagen 2014, 9.

[43] Rolf H. Weber/Ulrike I. Heinrich, Anonymization, London 2012, 15 ff.

[44] Baeriswyl (Fn 2), 53.

[45] Baeriswyl (Fn 2), 54.

[46] Baeriswyl (Fn 2), 52.

[47] Artikel-29-Datenschutzgruppe (Fn 22), 13.

[48] Vgl. dazu Artikel-29-Datenschutzgruppe (Fn 22), 15 ff., 29.

nicht bedeuten, dass der Gesetzgeber zwingend in rascher Weise neue Regelungen zu erlassen hätte; festhalten lässt sich nur, dass die traditionelle Datenschutzgesetzgebung keine besonderen Instrumente zur Verfügung stellt.

[Rz 21] Auf europäischer Ebene hat die Artikel-29-Datenschutzgruppe der Europäischen Union im Jahr 2007 eine Stellungnahme zum Konzept der personenbezogenen Daten abgegeben.[49] Darauf aufbauend hat diese Expertengruppe die bereits mehrfach erwähnte Stellungnahme zu Big Data vom 10. April 2014 erlassen. Die Vielzahl der Überlegungen zu den technischen Verfahren scheinen zwar für die Praxis wertvoll zu sein, doch sind die Empfehlungen für Juristen insofern ernüchternd, weil sie fast keinen rechtlichen Inhalt aufweisen. Die Hinweise beziehen sich vielmehr auf die technische Geeignetheit der Verfahren sowie auf kontextuelle und allgemein auf der Technik basierende Elemente.

[Rz 22] Der Entwurf für eine EU-Datenschutzgrundverordnung[50] sieht vor, eine angepasste Bestimmung für Zweckänderungen zu erlassen. Danach soll eine Verarbeitung der Daten mit Personenbezug möglich sein, auch wenn keine Kompatibilität mit dem ursprünglichen Zweck der Datenerhebung besteht; es ist lediglich vorausgesetzt, dass eine vertragliche Grundlage oder ein anderer Rechtfertigungsgrund, wie z.B. eine Einwilligung, vorliegt.[51] Dieses Modell würde eine Abkehr vom derzeit geltenden Modell bedeuten, wonach eine Verarbeitung von personenbezogenen Daten nur erlaubt ist, wenn diese in einer mit dem ursprünglichen Zweck zu vereinbarenden Art und Weise stattfindet.[52] Die Artikel-29-Datenschutzgruppe steht dem Vorschlag in der EU-Datenschutzgrundverordnung kritisch gegenüber und fordert die Streichung der Regelung.[53]

[Rz 23] In England gibt es verschiedene Richtlinien, welche vom lokalen Information Commissioner's Office (ICO) erarbeitet worden sind. Im Jahr 2007 hat die ICO in Reaktion auf die Stellungnahme der Artikel-29-Datenschutzgruppe der Europäischen Union eine erste «Guidance» erlassen.[54] Vor zwei Jahren hat dieselbe Behörde zuhanden der Unternehmen einen Code erarbeitet, welcher inhaltlich mit der Stellungnahme 5/2014 der Artikel-29-Datenschutzgruppe vergleichbar ist.[55]Weil es sich bei diesen Richtlinien nicht um ein Gesetz handelt, ist die Nichtbeachtung dieser Codes durch den Staat jedoch nicht sanktionierbar.[56]

---

[49] Artikel-29-Datenschutzgruppe der Europäischen Union, Stellungnahme 4/2007 zum Begriff «personenbezogene Daten», angenommen am 20. Juni 2007, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_de.pdf.

[50] Entwurf der EU-Datenschutzverordnung, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:DE:PDF.

[51] Vgl. dazu Art. 6 Abs. 4 sowie für die Rechtefertigungsgründe Art. 6 Abs. 1 lit. a-e des Entwurfs der EU-Datenschutzverordnung.

[52] Art. 6 Abs. lit. b der EG-Richtlinie 95/46/EG zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr, http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:31995L0046&from=de.

[53] Artikel-29-Datenschutzgruppe der Europäischen Union, Opinion 03/2013 on purpose limitation, angenommen am 2. April 2013, 41, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf; für weitere Hinweise vgl. Peter Katko/Ayda Babaei-Beigi, Accountability statt Einwilligung?, in: MMR 6/2014, 363.

[54] ICO, «Technical guidance note on determining what is personal data», https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf.

[55] ICO, «Anonymisation: managing data protection risk code of practice», https://ico.org.uk/media/1061/anonymisation-code.pdf.

[56] Eingehender dazu Saxby/Knight/Pearce (Fn 42), 14 ff.

3 «Management» von Big Data und Lösungsansätze

3.1 Big Data Governance

[Rz 24] Mit Bezug auf die Datenverwaltung und die Datenprozessierung scheint es unumgänglich, im Bereich von Big Data eine sachgerechte Governance einzuführen. Diese kann verschiedene Schritte betreffen, wie z.B. die Risikoanalyse, die Festlegung der Strategie, die Festlegung der eigentlichen Bearbeitungspolitik sowie prozessuale Verfahren, die dem Datenschutzumfeld gerecht werden:[57]

- *Risikoanalyse:* Im Rahmen der Risikoanalyse hat das Big Data Management insbesondere die Quellen der Daten zu eruieren. Ferner ist zu analysieren, für welche Zwecke die Organisation die Daten benötigt. Es ist auch zu beurteilen, ob sich die Big Data Analyse ausserhalb von vertrags- oder lizenzrechtlichen Bestimmungen bewegt, zu welchen sich das Unternehmen verpflichtet hat. Schliesslich sind die Erkenntnisse in einem Massnahmenplan festzuhalten.
- *Festlegung der Strategie:* Um die Strategie festlegen zu können, ist die Risikoanalyse als Grundlage heranzuziehen. Die Strategie sollte rationale Ziele und eine gute Governance für Big Data vorsehen; sodann sind die rechtlichen Risiken zu würdigen und die weiteren für eine Strategie notwendigen Punkte zu berücksichtigen.
- *Festlegung der Bearbeitungspolitik:* Im Rahmen der Bearbeitungspolitik ist ein detaillierter Projektplan zu erstellen. Dieser Projektplan sollte die Tools nennen, die für eine sachgemässe Anonymisierung der relevanten personenbezogenen Daten sorgen können. Darauf aufbauend ist festzulegen, wie die gewählten Verfahren implementiert werden können.
- *Einführung adäquater Verfahren:* In diesem vierten Stadium sind detaillierte Prozesse und Verfahren zu erarbeiten, welche den Umgang mit Big Data festhalten. Unternehmen wie auch der Staat müssen dafür sorgen, die für ihre Umgebung gewählten Richtlinien tatsächlich zu implementieren, einschliesslich der Einführung von (internen) Beschwerdeverfahren.

[Rz 25] Um die Risiken einer De-Anonymisierung von anonymisierten Daten zu senken, ist es somit unabdingbar, adäquate Massnahmen zu ergreifen. Zentral erscheinen insbesondere rechtliche, organisatorische und technische Sicherheitsmassnahmen.[58] Wie bereits im Rahmen der Risikoanalyse gezeigt, spielen rechtliche Abklärungen im Rahmen des Big Data Managements eine herausragende Rolle. In diesem Kontext sind u.a. die Herausforderungen bei der Durchsetzung von Datennutzungsrichtlinien im Falle von auftretenden Risiken der De-Anonymisierung im Auge zu behalten.[59]

[Rz 26] Die unterschiedlichen Niveaus, auf welchen sich eine solche Big Data Governance auszuwirken vermag, sind anhand der vorhandenen Umgebung zu identifizieren. Zudem sind folgende Fragestellungen zu beantworten: Welche Massnahmen erscheinen im Bereich Plattform/Infrastruktur als unvermeidlich und was gilt es im Kontext der Informationsarchitektur zu erledigen? In welcher Art und Weise ist es möglich, die rechtlichen Vorgaben einzuhalten und welche Grundsätze sind im Bereich Informationsmanagement/Informationssicherheit zu beachten?[60]

---

57 Vgl. vertiefend Kemp (Fn 19), 490.

58 Vgl. Saxby/Knight/Pearce (Fn 42), 23 ff.

59 Saxby/Knight/Pearce (Fn 42), 23.

60 Dazu vgl. Kemp (Fn 19), 486 f.; ebenfalls zu Big Data Governance vgl. Reinhard Riedl, Welchen Regulierungsbedarf schaffen die Paradigmenwechsel von Big Data?, in: Jusletter IT 21. Mai 2015.

[Rz 27] Eine gute Governance in den Unternehmen erscheint als notwendig, um dem Big Data Phänomen und seinen Risiken Herr zu werden. Zusätzlich sind aber weitere Massnahmen zu bedenken, welche zu einer Risikominderung bei der Datenbearbeitung beitragen können.

3.2 Risikominderung bei der Datenbearbeitung

[Rz 28] Im Kontext von Big Data Analytics ist das Risiko zu minimieren, von anonymisierten Daten mögliche Rückschlüsse auf Personen ziehen zu können. Diesbezüglich sind gesetzliche Regelungen denkbar, die unterschiedliche Verwertungsregeln statuieren: Möglich wäre die Pflicht des Datenbearbeiters, Daten sofort zu löschen, sobald erstellt ist, dass einmal anonymisierte Daten wieder Rückschlüsse auf Personen zulassen. Zudem ist auch die Frage aufzuwerfen, inwieweit im Kontext der Verwertung eine neue (nachträgliche) Einwilligung des Datenherrn einzuholen ist.[61]

[Rz 29] Weil die Bekanntgabe anonymisierter Daten keine Beschränkung von Verwertung und Auswertung kennt, wäre es möglich, die Bekanntgabe von anonymisierten Daten im Rahmen eines umfassenden Datenrechts zu regulieren.[62] Weiter wäre auch denkbar, ein grundsätzliches Verbot der Anonymisierung von gewissen Daten ins Auge zu fassen; dieser Lösungsansatz hätte zur Folge, dass die Anonymisierung nicht mehr aus dem Anwendungsbereich des DSG fallen könnte. Dieses mögliche Konzept scheint jedoch insofern als wenig überzeugend, weil der Schwerpunkt auf der Verbesserung von technischen Verfahren gelegt werden sollte, um eine mögliche De-Anonymisierung auszuschliessen.

[Rz 30] Im privatrechtlichen Bereich setzt Big Data die vorhandene Einwilligung als Rechtfertigungsgrund voraus, sofern mit einer De-Anonymisierung zu rechnen ist.[63] Diesbezüglich ist zu analysieren, was der Rechtfertigungsgrund der Einwilligung (Art. 13 DSG) eigentlich bedeutet. Genuin-juristisch setzt die Einwilligung voraus, dass das Individuum sich bewusst ist, wozu die Zustimmung erfolgt; eine Einwilligung hat also transparent zu erfolgen.[64] Bei Big Data ist diese Anforderung wohl in aller Regel nicht erfüllt.

[Rz 31] Die Problematik rund um die Einwilligung zeigt sich exemplarisch anhand verschiedenster Websites von Anbietern, die ihre Allgemeinen Geschäftsbedingungen (AGB) und Datenschutzerklärungen zur Verfügung stellen; diese Dokumente sind oft derart unübersichtlich gestaltet, dass das einwilligende Individuum kaum wissen kann, wozu es einwilligt. In der Praxis wird der Konsument zudem seine Einwilligung oft geben, ohne die entsprechenden Erklärungen genau gelesen zu haben. Somit ist eine frühe Einwilligung kein gutes Mittel, um das Einverständnis des entsprechenden Datensubjekts zu begründen; wenn überhaupt müsste die Einwilligung gestaffelt erfolgen.[65] Die Individuen sollten also mehrmals die Möglichkeit haben, ihre Einwilligung (transparent) erteilen zu können. Auch wenn dies in der Praxis umständlich sein mag, scheint ein solches Vorgehen bei Beibehaltung dieses Rechtfertigungsgrundes das einzige Mittel zu sein, eine «echte» Einwilligung annehmen zu können.

---

61 Eingehender dazu Baeriswyl (Fn 2), 55 f.

62 Baeriswyl (Fn 2), 56 f.

63 Baeriswyl (Fn 2), 56; Weber (Fn 5), N 25 ff.

64 Weber (Fn 5), N 27.

65 Vgl. dazu Rolf H. Weber, E-Commerce und Recht, 2. Aufl., Zürich 2010, 352; Bruno Baeriswyl, «Soziale Netzwerke» — Taktgeber für die Reform des Datenschutzrechts, in: Weber/Thouvenin (Hrsg.), Neuer Regulierungsschub im Datenschutzrecht?, ZIK Band 54, Zürich 2012, 93, 100 f.; Weber (Fn 5), N 26.

[Rz 32] Die Artikel-29-Datenschutzgruppe der Europäischen Union hat in einem Papier aus dem Jahre 2013 festgehalten, dass im Kontext von Big Data eine spezifische Einwilligung verlangt werden müsste, wenn sich eine Analyse auf bestimmbare Personen bezieht. Zudem empfiehlt die Expertengruppe, im Bereich von Online-Datenerhebungen leicht zugängliche und bedienungsfreundliche «Multi-Layer-Privacy-Notices» bereitzustellen.[66] Folglich gibt es einzelne Möglichkeiten und Bestrebungen, um die Beachtung datenschutzrechtlicher Regelungen zu verbessern. Gesamthaft gesehen erscheint es jedoch als sinnvoller, die organisatorischen Vorkehren der Unternehmen zu überprüfen; das diesbezügliche Stichwort lautet «Accountability».[67]

[Rz 33] Der Begriff Accountability ist kaum zu übersetzen; angesichts des lateinischen Wortes «*accomptare*» steht insbesondere die Rechenschaftspflicht der Unternehmer im Zentrum,[68] doch ist auch ihre Verantwortung angesprochen. Somit ist eine verstärkte Eigenkontrolle der Unternehmer, die als Datennutzer fungieren, zentral.[69] Weil die Analysemethoden im Kontext von Big Data geheim sind, umfasst Accountability auch die Pflicht zur Transparenz und zur Einhaltung der DSG-Grundsätze.[70] Zusätzlich sind gesetzliche Vorgaben denkbar, welche je nach Art der Verwendung eine bestimmte Datennutzung nur mit minimen oder gar ohne Datenschutzanforderungen zulassen könnten.[71]

4 Ausblick

[Rz 34] In Zukunft wird es wohl immer schwieriger werden, Personendaten und Sachdaten zu unterscheiden. Aus diesem Grunde wäre es sinnvoll, einzelne spezialgesetzliche Regelungen auszuarbeiten, um bestimmte Sachbereiche zu erfassen.[72] Im Bereich E-Health sind die Vorteile und Risiken von Big Data besonders ausgeprägt. Denkbar wären auch Vorschläge, welche auf die Frage abzielen, inwieweit die Bekanntgabe von anonymisierten Daten nicht möglich sein soll, weil das Risiko für eine De-Anonymisierung als sehr hoch eingeschätzt wird.

[Rz 35] Um das Informationsgefälle zwischen dem Datenbearbeiter und den Individuen zu verringern, scheint es zentral zu sein, in Form einer angepassten Einwilligung und einem implementierten Accountability-Standard grössere Transparenz zu schaffen. Sobald keine Ungleichgewichte mehr bestehen, ist eine adäquate Einwilligung auch eher möglich. Schliesslich ist für ein stärkeres «Identity Centric»-Konzept zu plädieren, damit Individuen (wieder) vermehrt die Kontrolle über ihre Daten erlangen. Sofern eine Kontrollmöglichkeit besteht, ist es auch einfacher, Daten selbst zu löschen.

[Rz 36] Die Datenschutzrevisionsüberlegungen, wie sie bisher angestellt worden sind, decken wohl nicht die Bedürfnisse und Notwendigkeiten ab, die sich in der Zukunft stellen; die Per-

---

66 Artikel-29-Datenschutzgruppe (Fn 53), 45 f. sowie Katko/Babaei-Beigi (Fn 53), 363.

67 Vgl. Weber (Fn 5), N 28 ff. und Katko/Babaei-Beigi (Fn 53), 360 ff.

68 Rolf H. Weber, Accountability in the Internet of Things, Computer Law & Security Review 27, 2011, 133 f.; Weber (Fn 5), N 28.

69 Mayer-Schönberger/Cukier (Fn 8), 177, 193.

70 Vgl. dazu Weber (Fn 5), N 16 f., 28.

71 Katko/Babaei-Beigi (Fn 53), 363.

72 Dies könnte z.B. im Gesundheitsbereich (E-Health) eine sachgemässe Lösung darstellen; vgl. bezüglich der Aktualität von E-Health die Entwicklungen zum elektronischen Patientendossier, http://www.e-health-suisse.ch/umsetzung/00135/00218/00256/index.html?lang=de.

spektive ist somit zu verbreitern. Insbesondere dürfte es auch im Interesse der Unternehmen sein, mehr Ressourcen für die Sicherheit der Daten zu investieren, um den Ausschluss der De-Anonymisierung einmal anonymisierter Daten zu gewährleisten. In Zukunft könnte Datenschutz zu einem Reputationslabel für Unternehmen werden. Nicht auszuschliessen ist gar ein Konditionenwettbewerb zwischen den Unternehmen, dessen Parameter die Höhe des Datenschutzniveaus festlegen.

[Rz 37] Sollte ein solcher Wettbewerb stattfinden, könnten sich die bestehenden Probleme bis zu einem gewissen Grad von selbst erübrigen. Tatsache ist jedenfalls, dass die Einwilligung, wie sie heute in Art. 13 DSG vorgesehen ist, sich nicht mehr als geeignet erweist, die Big Data-Sachverhalte aus dem 21. Jahrhundert sachgemäss abzudecken.

Prof. Dr. iur. Rolf H. Weber ist Ordinarius für Privat-, Wirtschafts- und Europarecht an der Universität Zürich, Visiting Professor an der Hong Kong University und praktizierender Rechtsanwalt in Zürich.
MLaw Dominic Oertly ist Assistent und Doktorand am Lehrstuhl von Prof. Dr. iur. Rolf H. Weber.